CESifo WORKING PAPERS

11671
2025

January 2025

Pattern, Perception, and Performance: Tripartite Phrases in Academic Paper Titles

Lutz Bornmann, Klaus Wohlrabe

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11671*

Pattern, Perception, and Performance: Tripartite Phrases in Academic Paper Titles

Abstract

This study examines how tripartite phrases in academic paper titles affect citation counts. Tripartite phrases consist of three interconnected elements separated by commas and conjunctions such as pattern, perception, and performance. Analyzing comprehensive datasets from economics (235,330 articles) and medicine and life sciences (93,713 articles), we found that papers with titles including tripartite phrases receive significantly more citations. Papers with tripartite phrases receive on average 3.5 additional citations in economics and 32 additional citations in medicine and life sciences compared to those without, controlling for paper characteristics, journal characteristics, and publication timing. For medical and life sciences papers, this effect persists when controlling for expert-assessed paper quality. The relative share of tripartite titles among the titles of all papers was stable over time at approximately 9% in economics and 4% in medicine and life sciences, suggesting an established stylistic convention.

JEL-Codes: A100.

Keywords: tripartite phrases, academic writing, bibliometrics, citation analysis, paper titles.

Lutz Bornmann
Science Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8
Germany – 80539 Munich
lutz.bornmann@gv.mpg.de

*Klaus Wohlrabe**
ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5
Germany – 81679 Munich
wohlrabe@ifo.de

*corresponding author

January 31, 2025

1 Introduction

In the modern science system, two instruments are usually used for research evaluation purposes: peer review and bibliometrics. Whereas peer review has been applied since the beginning of the modern science system, bibliometrics has been introduced in the 1950s and applied in research evaluation processes since the 1970s. Whereas publication counts are used to measure research output, citation counts are used to measure research impact whereas impact is seen as one part of quality (besides accuracy and importance). Today, bibliometrics plays an enormous role in research evaluation. For mathematics, Hulek and Teschke (2023) indicate that "various parameters are used to evaluate research performance, with bibliometric data playing an important role in (almost) all evaluations" (p. 36). The frequent use may be (partially) driven by "practicality considerations: Limited resources and a lack of incentives for committee work may make the use of such easily obtainable measures of academic productivity appear quite attractive" (Etzel et al., 2024). Other reasons for using bibliometrics are science wide application, global coverage and a low gaming potential (Ioannidis and Maniadis, 2023).

One of the greatest problems with the use of citation counts for research evaluation is its dependency on factors other than research quality such as the length of papers (all things being equal, longer papers tend to receive more citations). Studies suggest that the presentation of research – particularly specific paper titles – may attract not only readers but also citing authors. Recent bibliometric research suggests that title construction may significantly influence an article's visibility and subsequent citations, although the relationship is complex and may depend on disciplinary styles. Previous research on title construction and citation impact has yielded mixed results across disciplines. While some studies found that engaging titles correlate with higher citation counts, others suggest that attempts at crafting "amusing" titles may actually reduce citations (Sagi and Yechiam, 2008; Nair and Gibbert, 2016; Nazarovets and Silva, 2024). In this study, we continue this line of research by investigating whether tripartite phrases in paper titles have a significant effect on citations in two disciplines.

Title construction involving tripartite phrases follows the classical principle of "omne trium perfectum" – everything that comes in threes is perfect. This "rule of three" has been foundational in rhetoric since antiquity, exemplified by Julius Caesar's "veni, vidi, vici." (Julius Caesar was a Roman statesman and general). Cognitive science research suggests that information presented in small groups is more readily processed, with groups of three being particularly effective for human memory and comprehension. For empirical evidence on human cognitive processing of small information groups, we recommend the paper by Cowan (2001).

In academic writing, tripartite phrases in titles may serve multiple functions. They enhance clarity by decomposing complex ideas into interconnected components, create memorable rhythmic patterns, and efficiently communicate multiple research aspects within space constraints. For instance, titles like "Market size, trade, and productivity" (Melitz and Ottaviano, 2008) demonstrate how three interrelated concepts can be effectively linked. This structure may be particularly valuable in today's competitive academic landscape, where paper visibility and memorability significantly may influence citation patterns.

This study is based on comprehensive datasets from economics as well as medical and life sciences research. We examine whether papers with tripartite phrases in titles receive more citations than papers without such titles, controlling for confounding factors such as paper age, paper length, number of authors, and journal characteristics. Our empirical strategy employed multiple regression analyses – including OLS and negative binomial regression analyses – with journal and time fixed effects to account for disciplinary and time-specific citation practices. Since citation counts depend on the papers' quality, we controlled for the quality in some models to focus on factors in the analyses that are not related to quality. Our study was intended thus to contribute to bibliometric research on factors influencing citations (besides research quality). Since our empirical results suggest that structural elements of academic writing (tripartite phrases in titles) play a meaningful role in the citation process (independent of research quality), our findings have practical implications for researchers seeking to maximize their work's visibility and for the interpretation of results from citation analyses in research evaluation.

The remainder of this paper proceeds as follows: Section 2 reviews relevant literature for our research question. Section 3 describes our data by presenting some descriptive statistics. In this section, we also outline the empirical model. In Section 4, we present the empirical results. In Section 5, we summarize and give some outlook for future research.

2 Literature review

The sociology of science studies can be divided into three areas: (1) the study of institutions (Merton, 1973), (2) the study of procedures for discovering information (Zuckerman, 1988, p. 531), and (3) the study of the body of certified knowledge (Zuckerman, 1988, p. 531). Citation analysis falls into the third area, by where it examines the integration of citation links within the body of certified knowledge. The Social Systems Citation Theory (SSCT) proposed by Tahamtan and Bornmann (2022) offers a framework to conceptualize and explain processes and phenomena within this body, which is composed of publications and their citation links (Sjögårde, 2023; Sjögårde and Ahlgren, 2024). SSCT focuses on communication elements, such as publications and citations, as the fundamental building blocks of this (autopoietic) social science system. The theory suggests that processes and phenomena within this social science system can be understood through these building blocks that can be influenced by human elements such as the perception of certain paper titles.

Based on earlier reviews by Cronin (1984) and Liu (1993), Bornmann and Daniel (2008), Tahamtan and Bornmann (2019), and Kousha and Thelwall (2024) presented narrative reviews of studies that have investigated reasons for citing publications in empirical literature from the early 1960s. These literature overviews conclude that papers may be cited for various scientific and non-scientific reasons (and not only one dominant reason such as the quality of the cited publication). A similar conclusion comes from a meta-analysis (Glass, 1976) published by Mammola et al. (2022) including many factors potentially affecting citations. Whereas these summarizing papers present narrative overviews or meta-analyses of previous studies, Tahamtan and Bornmann (2018) provides a conceptual overview of the literature dealing with the process of citing. Conceptual means that the overview is "structured based

on core elements in the citation process: the context of the cited document, processes from selection to citation of documents, and the context of the citing document" (Tahamtan and Bornmann, 2018, p. 203). Reasons for citing that are derived from the context of the cited document include title, abstract, and keywords. Citing authors may select or reject a certain publication for citing based on its title, abstract, or keywords.

Numerous studies have highlighted that various characteristics of a publication's title, keywords, and abstract can influence how attractive and engaging they are to different readers. These characteristics encompass a wide range of elements, each contributing uniquely to the publication's appeal. One significant factor is the diversity and number of keywords present in the title or abstract, as demonstrated by studies such as those by Falagas et al. (2013) and Annalingam et al. (2014). The length of the title and abstract is another critical aspect, with research by Stremersch et al. (2015) showing that these elements' brevity or verbosity can significantly impact reader engagement. The presence of specific words in the publication's abstract, as explored by Ibanez et al. (2009), can enhance or detract from the publication's appeal. The format of the abstract itself – whether it is structured or unstructured – has also been examined, with findings by Lokker et al. (2008) that structured abstracts may be more appealing to certain audiences. The use of punctuation marks such as hyphens, commas, colons, and brackets in the title, studied by Buter and van Raan (2011), is another stylistic element that can affect reader perceptions and interest.

The inclusion of country names in the title, as identified by Jacques and Sebire (2010), and the type of title (whether it is a compound title, a question title, or a descriptive title) – investigated by Subotic and Mukherjee (2014) – also play roles in influencing how a publication is perceived. Reporting the study design in the title, as noted by Antoniou et al. (2015), can provide additional context that enhances the publication's appeal. Publications with shorter titles might attract more attention compared to those with longer titles, as suggested by Ayres and Vars (2000). This is because shorter titles can be quicker read and are easier to remember, thus potentially making them more appealing in fast-paced environments where readers are inundated with information. Another important feature is topicality, which refers to how relevant a publication is to the citing author's publication. According to Wang and

Soergel (1998), a publication's topicality can significantly influence its attractiveness. In general, publications that address hot topics tend to have higher topicality for citing authors. This higher relevance may translate to greater attention and more citations. Researchers typically prefer to write about current hot topics in their respective fields, and therefore, they frequently cite literature that is relevant to these topics (Fu and Aliferis, 2010; Gallivan, 2012).

3 Methods

3.1 Papers from medicine and life sciences including assessments by experts

We analyzed papers from the platform Faculty Opinions (FO, provided by H1, https://connect.h1.co), formerly F1000. FO is a post-publication appraisal and recommendation service in the fields of medicine and life sciences. Expert members, known as 'peers', rate papers on a three-level ordinal scale ('good', 'very good', 'excellent') to indicate their perceived quality. Note that neither low-quality nor ordinary-quality publications are rated as such; only contributions considered 'good' or better are recommended for inclusion in the FO database. These recommendations are made publicly under the peers' names within the FO subscription service and usually come with a brief explanation of the publication's significance. FO data, due to its large-scale dataset on concise peer reviews, has been widely used in bibliometric and altmetric research, as detailed in Williams (2017). Since peer ratings may be the best way of assessing the quality of papers (Bornmann, 2011), this expert-based quality assessment provides us with an additional control variable that helps account for the inherent quality of papers. Since any peer can rate any paper, papers are rated independently by multiple peers. For these papers, the multiple ratings from different experts are combined into a total article score.

On November 2023, we received a dataset snapshot from H1 including 246,245 peer ratings of publications from their service. We excluded 282 records: the records were dissent ratings

that express disagreement with recommendations by other peers. These recommendations do not fit into the three-level quality scale usually used for assessing the papers. The paper recommendations are from 2001 until 2023; most of the recommendations are from 2012 with more than 16,000 recommendations. Since any peer can rate any paper, many papers in the snapshot are rated independently by multiple peers. On average, papers received 1.2 recommendations, and 16% of the papers receiving more than one recommendation. The most assigned rating was 'good' (51%), followed by 'very good' (39%), and 'excellent' (10%). Since many papers received more than one rating, we computed the total article score by summing up the ratings for this study. Using the publication date and the recommendation date, we can determine the average time between publication and recommendation date which is 222 days with a standard deviation of 765 days. About 14% of the recommendations have been published before the recorded official publication date (possibly based on preprints). While the short average time between publication and recommendation suggests that FO ratings are scarcely influenced by citation impact information, it may be possible that citing authors are partially informed and influenced by FO ratings.

3.2 Bibliometric data

Bibliographic and citation data for the papers used in this study have been derived from an in-house database at the Max Planck Society that is based on the Web of Science (WoS, Clarivate): publication year, title, first page, last page, number of authors, journal title, and document type. We have also derived citation counts for three periods: (i) between publication year and 2023 (total citations), and for a fixed citation window of three (ii) and five (iii) years.

The first dataset for this study including papers published in medicine and life sciences have been produced by matching the FO data from H1 to the papers in the in-house database using the DOI. The second dataset including the economic papers has been produced based on WoS subject categories: we considered all papers from the categories with 'economics' in the category name published between 2006 and 2019. We selected these publication years –

not only for the economics, but also the medicine and life sciences dataset – because of two reasons: (1) For earlier publication years (2001-2005), only few recommended papers are in the dataset from H1. (2) To reliably measure citation impact, the citation window should not be too short. With at least five years for the most recent publication year (2019), the citation window is sufficient long in this study for all included publications in both disciplines. The original dataset in economics comprised 414,836 and in medicine and life science 146,174 articles.

We have made the following additional adjustments to both disciplinary initial datasets. We kept only those with the document type 'article' (since one can expect different citation impact for different types). Thus, we did not consider (book) reviews, proceedings, editorials etc. We restricted the analysis to papers with available page information and required a minimum of 20 papers per journal to enable reliable estimations of journal fixed effects. The final samples consist of 235,330 articles from 398 economics journals and 93,713 articles from 624 medicine and life sciences journals.

3.3 Identification of tripartite phrases in titles

To identify tripartite phrases in the paper titles, we employed three complementary algorithms that analyzed the grammatical structure and connectors within titles. Each algorithm evaluates potential tripartite structures through different approaches: (1) counting commas and checking for 'and' conjunctions, (2) analyzing part segmentation through split operations, and (3) using regular expressions to identify list-like structures. A title is classified as having a tripartite structure only if all three algorithms agree, ensuring high precision in identification. Specifically, a title must either contain at least two commas plus an 'and' connector, or one comma with a subsequent 'and' that joins the final elements. This conservative approach, requiring unanimous agreement across all three methods, minimizes false positives while reliably identifying clear tripartite structures in titles.

To assess the reliability of our automated identification approach, we conducted a detailed validation study using a random sample of 1,000 articles from our economics dataset. The

automated algorithm identified 103 articles containing tripartite phrases. Through careful hand-coding by the authors, we confirmed 98 of these identifications (95% accuracy). Among the five misclassified cases, three involved listings of countries rather than true tripartite structures. We also examined 12 cases where our algorithmic approach yielded inconclusive results (identified by at least one but not all three algorithms). Hand-coding revealed that five of these actually contained tripartite phrases. Additionally, manual inspection identified one article with a tripartite phrase that was not detected by any of our algorithms. Despite these few misclassified cases, our results suggest that our automated approach achieves high precision in identifying tripartite phrases, with a false positive rate of only 5%. The algorithm's main limitation appears to be in distinguishing between true tripartite structures and simple listings, i.e. listings of countries. While it may miss some edge cases (as evidenced by the five confirmed tripartite phrases among inconclusive cases and one undetected case), its high accuracy rate makes it applicable in our investigation.

To illustrate the variety and structure of tripartite phrases in academic titles, Table 1 presents some representative examples from both economics and the medical and life sciences paper sets. In economics, the examples demonstrate how tripartite phrases effectively combine related theoretical concepts (e.g., "Envy, inequality and fertility") or methodological elements (e.g., "Market concentration, collusion and social welfare in Mexico: A methodological update"). The examples from medical and life sciences show how tripartite phrases are often used to describe clinical processes (e.g., "Hepatitis C in pregnancy: screening, treatment, and management") or experimental procedures (e.g., "RNA-Catalyzed polymerization of deoxyribose, threose, and arabinose nucleic acids"). These examples from both disciplines highlight how tripartite structures can efficiently communicate multiple interconnected aspects of research while maintaining clarity and conciseness within title length constraints.

Our analysis identified 21,434 articles with tripartite phrases in economics (9.2% of the sample) and 4,088 articles in medicine and life sciences (4.3% of the sample). Table 2 presents descriptive statistics for the citation impact of papers published in both disciplines. The results show that papers with tripartite phrases in economics received on average 29 total citations compared to 24 for those papers without these phrases. The most cited paper

Table 1: Title examples containing tripartite phrases

Economics
Envy, inequality and fertility
Informed trade, uninformed trade and stock price delay
Housing, adjustment costs, and macro dynamics
Trade, development, and poverty-induced comparative advantage
Market concentration, collusion and social welfare in Mexico: A methodological update
Medicine and life sciences
Hepatitis C in pregnancy: screening, treatment, and management
RNA-Catalyzed polymerization of deoxyribose, threose, and arabinose nucleic acids
Hospitalization, surgery, and incident dementia
Red meat, dairy, and insulin sensitivity: a randomized crossover intervention study
The B73 maize genome: complexity, diversity, and dynamics

with a tripartite phrase is Dohmen et al. (2011) ("Individual risk attitudes: Measurement, determinants, and behavioral consequences") with 1,659 total citations. The counterpart paper (the most cited paper) without a tripartite phrase is Pesaran (2007) ("A simple panel unit root test in the presence of cross-section dependence") which collected 4,800 citations.

In medicine and life sciences, articles with tripartite phrases accumulate on average 175 total citations versus 168 for those without these phrases. The article with the most total citations containing a tripartite phrase is Topalian et al. (2012) ("Safety, activity, and immune correlates of anti-PD-1 antibody in cancer") with 8,805 citations. The paper by Schneider et al. (2012) received 39,336 citations without a tripartite phrase ("NIH Image to ImageJ: 25 years of image analysis").

The pattern of higher citation counts for tripartite phrase's papers persists across shorter citation windows (instead of the total citation counts): citations within three- and five-year windows after publication (see Table 2). Statistical t-tests confirm that the differences in mean citations between papers with and without tripartite phrases are statistically significant at the 1% level for all citation windows in both disciplines. Since these descriptive differences do not account for potential confounding factors (research quality and factors influencing citations beyond tripartite phases in the titles), we also calculated regression models.

The citation patterns of economics as well as medicine and life sciences display notable

differences in scale and distribution. Medical and life sciences papers generally received substantially more total citations than economics papers, with average total citation counts roughly six times higher (169 versus 25). This difference is consistent across both tripartite and non-tripartite phrases' papers and persists in shorter citation windows. The absolute citation impact advantage for tripartite phrases' papers also appears larger in medicine and life sciences, although the relative difference is more pronounced in economics.

Figure 1 illustrates the temporal development of tripartite phrases usage in papers from both disciplines over the period 2006–2019. The absolute numbers show distinct patterns across disciplines, while the relative shares remain remarkably stable. In economics, we observe a steady increase in the absolute numbers of tripartite phrases' papers from approximately 900 papers in 2006 to around 1,800 papers in 2012. After 2012, the numbers stabilize at this higher level through 2019. This pattern reflects the general expansion of published research in economics during the period. In medicine and life sciences, we observe an initial increase from 2006 to a peak around 2012, followed by an apparent decline. However, this decline in absolute numbers is attributable to the reduced coverage of publications in our sample rather than a decrease in the use of tripartite phrases, as evidenced by the stable relative share. Notably, despite these different temporal patterns in absolute numbers, the relative share of tripartite phrases' papers remains consistently stable at approximately 9% for economics and 4% for medicine and life sciences throughout our observation period. This stability suggests that tripartite phrasing represents an established stylistic convention in academic writing rather than a passing temporary trend.

3.4 Statistical model

Following Wohlrabe and Bornmann (2022), we tested whether tripartite phrases in titles have a statistically significant effect on citation impact by estimating the following model:

$$cit_i = \alpha + \beta \cdot Tripartite_i + \gamma \cdot X_i + \chi_j + year_t \tag{1}$$

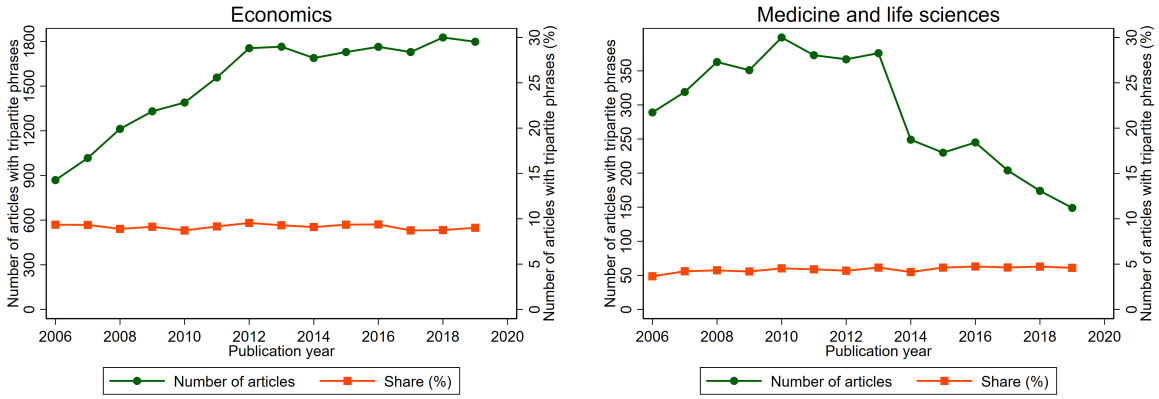
where β is our coefficient of interest. The variable $Tripartite_i$ is a dummy variable that takes

Table 2: Descriptive citations statistics

Category	N	Mean	Median	Minimum	Maximum	Standard deviation	
Economics							
Citations 3 year window							
Total	235,330	3	2	0	632	6	
Tripartite = 1	21,434	4	2	0	152	6	
Tripartite = 0	213,896	3	2	0	632	6	
p-value							0.00
Citations 5 year window							
Total	235,330	8	4	0	1,405	16	
Tripartite = 1	21,434	9	4	0	492	17	
Tripartite = 0	213,896	8	4	0	1,405	15	
p-value							0.00
Total citations							
Total	235,330	25	9	0	4,800	61	
Tripartite = 1	21,434	29	11	0	1,659	64	
Tripartite = 0	213,896	24	9	0	4,800	60	
p-value							0.00
Medical and life sciences							
Citations 3 year window							
Total	93,713	32	18	0	4,006	61	
Tripartite = 1	4,088	33	18	0	1,736	64	
Tripartite = 0	89,625	32	18	0	4,006	61	
p-value							0.00
Citations 5 year window							
Total	93,713	65	35	0	7,064	128	
Tripartite = 1	4,088	68	36	0	2,987	141	
Tripartite = 0	89,625	65	35	0	7,064	127	
p-value							0.00
Total citations							
Total	93,713	169	81	0	39,336	396	
Tripartite = 1	4,088	175	85	0	8,805	352	
Tripartite = 0	89,625	168	81	0	39,336	398	
p-value							0.00

Notes: The p-values of all t tests are statistically significant at the 1% level.

Figure 1: Development of articles with tripartite phrases in titles over time



the value 1 if the title of article i contains a tripartite phrase and 0 otherwise. We include several control variables (X_i) that previous research has identified as potential determinants of citation counts (Tahamtan and Bornmann, 2018). These include the number of pages (Ball, 2008), article age (Parolo et al., 2015), title length (Bramoullé and Ductor, 2018), occurrence of non-alphanumeric characters like questions or exclamation marks (Gnewuch and Wohlrabe, 2017), and the number of authors (Larivière et al., 2015). To account for journal-specific quality differences and time-dependent citation practices, we include journal fixed effects (χ_j) and time fixed effects ($year_t$), respectively. We considered only those determinants from the many determinants investigated in previous research with available data in our disciplinary datasets (and in-house database).

To ensure robustness of our results and to address the typically skewed distribution of citation counts, we followed the estimation strategy applied by Wohlrabe and Bornmann (2022) and estimated four different regression specifications:

1. A basic OLS regression model using raw citation counts as the dependent variable
2. A negative binomial regression (NBR) model that explicitly accounts for citations being count data
3. An OLS regression using the natural logarithm of citations as the dependent variable
4. An OLS regression using the inverse hyperbolic sine (IHS) transformation of citations.

This transformation, originally proposed by Burbidge et al. (1988) and recently applied to citation counts (c) by Card and DellaVigna (2020), is defined as:

$$\text{asinh}(c) = \ln(c + \sqrt{1 + c^2}) \tag{2}$$

where for $c \geq 2$, $\text{asinh}(c) = \ln(c) + \ln(2)$, but $\text{asinh}(0) = 0$. This transformation has properties similar to a logarithmic transformation but is well-defined at zero. The well-definition at zero makes the model particularly suitable for citation data that includes non-cited papers.

Our multi-model approach allowed us to verify whether the findings are not driven by any particular specification choice or treatment of the dependent variable. If tripartite phrases in titles indeed enhance a paper's visibility and citations, we would expect β to be positive and statistically significant across the different specifications.

Table 3 reports descriptive statistics for the explanatory variables. The age of an article has been calculated by 2024 − publication year + 1. For the variable non-alphanumeric characters in titles we considered [.; :!?]. We did not consider the comma for computing the variable, as it is closely related to our definition of tripartite phrases. For the title length variable, we counted the number of characters.

The descriptive statistics reveal notable differences between economics as well as medicine and life sciences papers across several dimensions. Economics papers in our sample have a mean age of 12 years, while medical and life sciences papers are slightly older with a mean age of 14 years. Paper length varies substantially between disciplines, with economics papers being generally longer (mean: 18 pages) compared to medical and life sciences papers (mean: 9 pages). Title construction also differs markedly between the disciplines: medical and life sciences papers tend to have longer titles with a mean of 104 characters versus 78 characters in economics. The use of non-alphanumeric characters is more prevalent in economics paper titles (mean: 0.5) than in medical and life sciences paper titles (mean: 0.2). Perhaps the most striking disciplinary difference appears in authorship patterns, with medical and life sciences papers having substantially more authors on average (mean: 9) compared to economics papers

(mean: 2). These characteristics possibly reflect different research and publication practices across both disciplines, which we accounted for in our subsequent regression analyses.

Table 3: Descriptive statistics for explanatory variables

	Mean	Median	Minimum	Maximum	Standard deviation
Economics					
Article age	12	12	6	19	4
Number of pages	18	17	1	581	10
Title length	78	75	4	353	29
Non-alphanumeric characters	0.5	0.0	0.0	1.0	0.5
Number of authors	2	2	1	282	1
Medicine and life sciences					
Article age	14	14	6	19	4
Number of pages	9	9	1	1,762	8
Title length	104	99	8	782	35
Non-alphanumeric characters	0.2	0.0	0.0	1.0	0.4
Number of authors	9	7	1	3,389	28
Article quality	1.9	2	1	41	1.6

4 Results

Tables 4 and 5 present the results of the regression analyses examining the relationship between tripartite phrases in article titles and citation counts across economics as well as medicine and life sciences, respectively. The results consistently show that articles with tripartite phrases received statistically significantly more citations compared to those without, even after controlling for various article and author characteristics. The control variables in our models generally show that the expected relations between tripartite phrases in titles and citation counts have empirical evidence and are statistically significant across various specifications.

For economics papers (see Table 4), tripartite phrases in titles are associated with a positive and statistically significant increase in citations across all model specifications. The OLS estimates indicate that papers with tripartite phrases received on average about 3.5 additional citations compared to papers without such phrases. This effect remains robust when using alternative specifications including negative binomial regression and logarithmic

transformations of citation counts. As expected, the magnitude of the effect is smaller but statistically significant when examining shorter citation windows of three and five years, probably reflecting the naturally lower citation counts in these shorter time periods.

The impact of tripartite phrases appears even stronger for research in medicine and life sciences than in economics (Table 5). The OLS estimates suggest a substantial citation premium of approximately 32 additional citations for articles with tripartite phrases compared to those without. This effect persists across all model specifications, including the negative binomial model and alternative citation metrics. Similar to economics, the three year and five year citation windows consistently show positive effects, albeit of smaller magnitude due to the shorter accumulation period.

Table 6 extends our analysis of medical and life science papers by incorporating the FO article quality scores. In this analysis, the coefficient for tripartite phrases remains nearly identical and statistically significant after controlling for quality. The result suggests that the citation advantage associated with tripartite phrases may not be driven by the underlying paper quality. As expected, the article quality score itself shows a strong positive relationship with citations, confirming that expert-assessed quality predicts citation impact. The substantial shift in the constant term in Table 6 indicates that much of the baseline citation level is explained now by the article quality score. The stability of the tripartite coefficient after including this additional quality control strengthens our confidence in the relationship between tripartite phrases and citations, although we cannot claim causality due to potential unobserved factors influencing citations.

As a robustness check, we replicated our analysis excluding cases where the classification of tripartite phrases was ambiguous according to our identification algorithm (see Tables 7 to 9 in the Appendix). In economics, there were 2,939 unclear cases, and in medicine and life sciences 2,785 cases. The results remain virtually unchanged across all specifications and samples. For both economics as well as medicine and life sciences, the coefficients for tripartite phrases maintain their magnitude and statistical significance, and the relationships between citations and control variables, including the article quality scores, remain stable. This consistency suggests that our findings are not driven by potential misclassification of

title structures.

Table 4: Results of regression models based on papers from economics

	OLS Total citations	NBR Total citations	OLS Log citations	OLS IHS citations	OLS 3 year citations	OLS 5 year citations
Tripartite phrase	3.498***	0.131***	0.104***	0.115***	0.320***	0.988***
	(0.421)	(0.011)	(0.007)	(0.008)	(0.041)	(0.111)
Article age	2.142***	0.083***	0.062***	0.069***	-0.200***	-0.431***
	(0.077)	(0.002)	(0.001)	(0.001)	(0.006)	(0.015)
Number of pages	0.653***	0.024***	0.021***	0.024***	0.069***	0.184***
	(0.039)	(0.001)	(0.001)	(0.001)	(0.004)	(0.009)
Title length	-0.052***	-0.002***	-0.000***	-0.000***	-0.004***	-0.012***
	(0.005)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Non-alphanumeric characters	4.428***	0.175***	0.142***	0.159***	0.412***	1.208***
	(0.264)	(0.008)	(0.005)	(0.005)	(0.028)	(0.070)
Number of authors	2.310***	0.127***	0.093***	0.106***	0.384***	0.911***
	(0.463)	(0.003)	(0.018)	(0.021)	(0.078)	(0.186)
Constant	-36.144***	-0.497***	-0.321***	-0.196***	1.476***	2.051***
	(1.422)	(0.094)	(0.057)	(0.068)	(0.204)	(0.487)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Journal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	235,330	235,330	235,330	235,330	235,330	235,330

Notes: The table reports coefficients and corresponding standard errors in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 5: Results of regression models based on papers from medicine and life sciences

	OLS Total citations	NBR Total citations	OLS Log citations	OLS IHS citations	OLS 3 year citations	OLS 5 year citations
Tripartite phrase	32.306***	0.181***	0.179***	0.183***	3.157***	8.871***
	(5.225)	(0.021)	(0.014)	(0.014)	(0.882)	(1.988)
Article age	10.446***	0.092***	0.079***	0.081***	-0.615***	-1.030***
	(0.591)	(0.002)	(0.002)	(0.002)	(0.168)	(0.318)
Number of pages	2.987	0.044***	0.007*	0.007*	0.925	1.694
	(1.825)	(0.004)	(0.004)	(0.004)	(0.612)	(1.109)
Title length	-0.267***	-0.002***	-0.001***	-0.001***	-0.015	-0.045**
	(0.052)	(0.000)	(0.000)	(0.000)	(0.010)	(0.019)
Non-alphanumeric characters	32.715***	0.210***	0.149***	0.152***	3.851***	9.102***
	(5.016)	(0.013)	(0.009)	(0.009)	(0.481)	(1.125)
Number of authors	1.089***	0.008***	0.002***	0.002***	0.261***	0.557***
	(0.257)	(0.001)	(0.000)	(0.000)	(0.059)	(0.124)
Constant	-88.459***	2.525***	2.604***	3.226***	14.698**	26.502**
	(20.736)	(0.092)	(0.080)	(0.083)	(6.614)	(12.195)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Journal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	93,713	93,713	93,713	93,713	93,713	93,713

Notes: The table reports coefficients and corresponding standard errors in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

5 Discussion

Since the beginning of the use of citation counts in research evaluation, one is interested in the factors that are responsible for citation decisions of authors. Besides the quality of the cited papers, many other factors have been identified in countless empirical studies possibly influencing the citation of papers. In the interpretation of empirical results in research evaluation based on citation counts, it is important to consider that the counts are not only quality driven. In this study, we have investigated for the first time whether tripartite phrases in article titles are an additional factor influencing citations that is not quality related.

Our analyses revealed a robust positive relationship between tripartite phrases in titles and citation counts across both economics as well as medicine and life sciences. While the effect is present in both disciplines, its magnitude differs substantially: medical and life

Table 6: Results of regression models based on papers from medicine and life sciences including article quality scores

	OLS Total citations	NBR Total citations	OLS Log citations	OLS IHS citations	OLS 3 year citations	OLS 5 year citations
Tripartite phrase	32.294***	0.181***	0.179***	0.183***	3.155***	8.868***
	(5.172)	(0.020)	(0.014)	(0.014)	(0.876)	(1.973)
Article age	11.318***	0.093***	0.081***	0.084***	-0.491***	-0.772**
	(0.596)	(0.002)	(0.002)	(0.002)	(0.167)	(0.316)
Number of pages	2.922	0.042***	0.007*	0.007*	0.916	1.675
	(1.777)	(0.003)	(0.004)	(0.004)	(0.606)	(1.096)
Title length	-0.248***	-0.002***	-0.001***	-0.001***	-0.012	-0.039**
	(0.051)	(0.000)	(0.000)	(0.000)	(0.010)	(0.019)
Non-alphanumeric characters	30.176***	0.203***	0.143***	0.146***	3.489***	8.350***
	(5.020)	(0.013)	(0.009)	(0.009)	(0.481)	(1.125)
Number of authors	1.071***	0.008***	0.002***	0.002***	0.259***	0.551***
	(0.257)	(0.001)	(0.000)	(0.000)	(0.059)	(0.123)
Article quality	39.277***	0.099***	0.093***	0.094***	5.609***	11.630***
	(2.711)	(0.003)	(0.002)	(0.002)	(0.295)	(0.665)
Constant	-163.041***	2.364***	2.427***	3.047***	4.047	4.419
	(21.176)	(0.093)	(0.079)	(0.081)	(6.527)	(12.061)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Journal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	93,713	93,713	93,713	93,713	93,713	93,713

Notes: The table reports coefficients and corresponding standard errors in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

sciences papers with tripartite phrases in titles received about 32 additional citations compared to 3.5 additional citations in economics. This difference likely reflects the generally higher citation counts in medicine and life sciences rather than differential effectiveness of tripartite structures. The stability of our findings across multiple regression model specifications and the robustness of the findings to controlling for expert-assessed quality through FO article quality scores strengthens the confidence in this relationship, although we cannot claim causality.

Our findings have practical implications for researchers crafting paper titles. Tripartite phrases appear to enhance article visibility without sacrificing scholarly credibility. We found not only an effect of tripartite phrases in titles, but also an effect of the papers' quality on citation counts. The consistent relative share of tripartite phrases over time (around 9% in economics and 4% in medicine and life sciences) suggests an established stylistic convention

rather than a temporary trend.

Our study has some limitations: (1) While we controlled for various confounding factors and paper quality scores in our study, unobserved confounding variables might have still influenced both title phrases choices and citation success. (2) Our focus on two disciplines in this study, although providing useful contrasts, limits the generalizability of the empirical results and could be extended by considering other disciplines. (3) The relationship between tripartite phrases in titles and citations might vary by article type (e.g., theoretically-oriented versus empirically-oriented papers). Since an article type variable was not available in both datasets used in this study, we have not considered this factor in our analyses. Despite the three limitations, our findings contribute to the understanding of how structural elements of paper titles influence citation impact. The consistent positive relationship between tripartite phrases in titles and citations suggests that careful title construction may enhance article visibility – independent of the quality of the reported research in the article.

Several promising directions for future research emerge from our study (based on the reported limitations). First, further confounding variables could be included in regression models in future studies on the influence of tripartite phrases in titles on citation counts (that were not available in our datasets). For example, one could explore whether the relationship between tripartite phrases in titles and citation counts varies by document type (books, conference proceedings etc.) or article type (theoretically-oriented versus empirically-oriented papers). Second, extending the analyses to other disciplines (besides economics as well as medicine and life sciences) would help to establish whether the effectiveness of tripartite phrases in titles generalizes across a broad spectrum of disciplines. Disciplines like physics, biology, and computer science might show different patterns due to varying citation practices and writing conventions. For the analyses in other disciplines, it would be especially interesting to have datasets including article quality scores based on expert assessments. Third, future studies could explore how tripartite phrases are utilized across different languages and cultures. This would involve analyzing their prevalence, structural variations, and effectiveness in various cultural contexts, thereby providing a more global understanding of their communicative power. An example of a tripartite phrase in a German title is

"Sozioökonomischer Status, Mentoring und Chancengerechtigkeit: Thünen-Vorlesung 2022" (Falk et al., 2023).

Acknowledgements

Since our study is mainly based on data from the platform Faculty Opinions (FO, provided by H1, https://connect.h1.co), we thank H1 for providing us with the data. We also used ivsdb bibliometric data. The ivsdb data are developed in cooperation between the Administrative Headquarters of the Max Planck Society (MPG) and the Information Retrieval Service for the institutes of the Chemical-Physical-Technical (IVS-CPT) Section of the MPG. The data stem from a bibliometrics database based on data provided by the German Kompetenznetzwerk Bibliometrie (KB, Competence Network Bibliometrics, funded by BMBF via grant 01PQ17001, see: http://www.bibliometrie.info/). The data are derived from the Science Citation Index - Expanded (SCI-E), the Social Sciences Citation Index (SSCI), the Conference Proceedings Citation Index - Science (CPCI-S), the Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH), and the Arts and Humanities Citation Index (AHCI), provided by Clarivate.

References

Annalingam, A., Damayanthi, H., Jayawardena, R., and Ranasinghe, P. (2014). Determinants of the citation rate of medical research publications from a developing country. *SpringerPlus*, 3(1):140.

Antoniou, G. A., Antoniou, S. A., Georgakarakos, E. I., Sfyroeras, G. S., and Georgiadis, G. S. (2015). Bibliometric analysis of factors predicting increased citations in the vascular and endovascular literature. *Annals of Vascular Surgery*, 29(2):286–292.

Ayres, I. and Vars, F. E. (2000). Determinants of citations to articles in elite law reviews. *Journal of Legal Studies*, 29(1):427–450.

Ball, P. (2008). A longer paper gathers more citations. *Nature*, 455(7211):274–276.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45:199–245.

Bornmann, L. and Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1):45–80.

Bramoullé, Y. and Ductor, L. (2018). Title length. *Journal of Economic Behavior & Organization*, 150:311–324.

Burbidge, J. B., Magee, L., and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83(401):123–127.

Buter, R. K. and van Raan, A. F. J. (2011). Non-alphanumeric characters in titles of scientific publications: An analysis of their occurrence and correlation with citation impact. *Journal of Informetrics*, 5(4):608–617.

Card, D. and DellaVigna, S. (2020). What do editors maximize? Evidence from four economics journals. *Review of Economics and Statistics*, 102(1):195–217.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114.

Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. Taylor Graham, Oxford, UK.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Etzel, F., Seyffert-Müller, A., Schönbrodt, F. D., Kreuzer, L., Gärtner, A., Knischewski, P., and Leising, D. (2024). Inter-rater reliability in assessing the methodological quality of

research papers in psychology. https://osf.io/preprints/psyarxiv/4w7rb, Retrieved May 13, 2024.

Falagas, M. E., Zarkali, A., Karageorgopoulos, D. E., Bardakas, V., and Mavros, M. N. (2013). The impact of article length on the number of future citations: A bibliometric analysis of general medicine journals. *PLOS One*, 8(2):e49476.

Falk, A., Kosse, F., and Stötzer, L. S. (2023). Sozioökonomischer Status, Mentoring und Chancengerechtigkeit: Thünen-Vorlesung 2022. *Perspektiven der Wirtschaftspolitik*, 24(2):264–275.

Fu, L. D. and Aliferis, C. F. (2010). Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. *Scientometrics*, 85(1):257–270.

Gallivan, M. J. (2012). Analyzing citation impact of IS research by women and men: Do women have higher levels of research impact? In *Proceedings of the 50th annual conference on Computers and People Research (SIGMIS-CPR '12)*, pages 175–184, New York, NY, USA. Association for Computing Machinery.

Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher*, 5:3–8.

Gnewuch, M. and Wohlrabe, K. (2017). Title characteristics and citations in economics. *Scientometrics*, 110:1573–1578.

Hulek, K. and Teschke, O. (2023). How do mathematicians publish? Some trends. *European Mathematical Society Magazine*, 129:36–41.

Ibanez, A., Larranaga, P., and Bielza, C. (2009). Predicting citation count of bioinformatics papers within four years of publication. *Bioinformatics*, 25(24):3303–3309.

Ioannidis, J. P. A. and Maniadis, Z. (2023). In defense of quantitative metrics in researcher assessments. *PLOS Biology*, 21(12):e3002408.

Jacques, T. S. and Sebire, N. J. (2010). The impact of article titles on citation hits: An analysis of general and specialist medical journals. *JRSM Short Reports*, 1.

Kousha, K. and Thelwall, M. (2024). Factors associating with or predicting more cited or higher quality journal articles: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology*, 75(3):215–244.

Larivière, V., Gingras, Y., Sugimoto, C. R., and Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7):1323–1332.

Liu, M. (1993). Progress in documentation - the complexities of citation practice: A review of citation studies. *Journal of Documentation*, 49:370–408.

Lokker, C., McKibbon, K. A., McKinlay, R. J., Wilczynski, N. L., and Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: Retrospective cohort study. *British Medical Journal*, 336(7645):655–657.

Mammola, S., Piano, E., Doretto, A., Caprio, E., and Chamberlain, D. (2022). Measuring the influence of non-scientific features on citations. *Scientometrics*, 127(7):4123–4137.

Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75(1):295–316.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press, Chicago, IL, USA.

Nair, L. B. and Gibbert, M. (2016). What makes a 'good' title and (how) does it matter for citations? A review and general model of article title attributes in management science. *Scientometrics*, 107:1331–1359.

Nazarovets, S. and Silva, J. A. (2024). The incidence of 11 movie titles in the titles of Scopus-indexed papers. *Scientometrics*, 129(5):2911–2922.

Parolo, P. D. B., Pan, R. K., Ghosh, R., Huberman, B. A., Kaski, K., and Fortunato, S. (2015). Attention decay in science. *Journal of Informetrics*, 9(4):734–745.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2):265–312.

Sagi, I. and Yechiam, E. (2008). Amusing titles in scientific journals and article citation. *Journal of Information Science*, 34(5):680–687.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*, 9(7):671–675.

Sjögårde, P. (2023). *Mapping the structure of science through clustering in citation networks: Granularity, labeling and visualization*. PhD thesis, Karolinska Institutet, Stockholm, Sweden.

Sjögårde, P. and Ahlgren, P. (2024). Seed-based information retrieval in networks of research publications: Evaluation of direct citations, bibliographic coupling, co-citations, and pubmed-related article score. *Journal of the Association for Information Science and Technology*, 75(13):1453–1465.

Stremersch, S., Camacho, N., Vanneste, S., and Verniers, I. (2015). Unraveling scientific impact: Citation types in marketing journals. *International Journal of Research in Marketing*, 32(1):64–77.

Subotic, S. and Mukherjee, B. (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science*, 40(1):115–124.

Tahamtan, I. and Bornmann, L. (2018). Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics*, 12(1):203–216.

Tahamtan, I. and Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics*, 121(3):1635–1684.

Tahamtan, I. and Bornmann, L. (2022). The Social Systems Citation Theory (SSCT): A proposal to use the social systems theory for conceptualizing publications and their citation links. *Profesional de la Información*, 31(4):e310411.

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., Powderly, J. D., Carvajal, R. D., Sosman, J. A., Atkins, M. B., et al. (2012). Safety, activity, and immune correlates of anti–PD-1 antibody in cancer. *New England Journal of Medicine*, 366(26):2443–2454.

Wang, P. and Soergel, D. (1998). A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science*, 49(2):115–133.

Williams, A. E. (2017). F1000: An overview and evaluation. *Information and Learning Science*, 118(7/8):364–371.

Wohlrabe, K. and Bornmann, L. (2022). Alphabetized co-authorship in economics reconsidered. *Scientometrics*, 127(5):2173–2193.

Zuckerman, H. (1988). The sociology of science. In Smelser, N. J., editor, *Handbook of sociology*, pages 511–574. Sage Publications, Inc., Thousand Oaks, CA, US.

Appendix

Table 7: Results of regression models based on papers from economics - excluding unclear cases

	OLS Total citations	NBR Total citations	OLS Log citations	OLS IHS citations	OLS 3 year citations	OLS 5 year citations
Tripartite phrase	3.558***	0.134***	0.105***	0.117***	0.326***	1.006***
	(0.421)	(0.011)	(0.007)	(0.008)	(0.041)	(0.111)
Article age	2.141***	0.082***	0.062***	0.069***	-0.200***	-0.431***
	(0.077)	(0.002)	(0.001)	(0.001)	(0.006)	(0.016)
Number of pages	0.651***	0.024***	0.021***	0.024***	0.068***	0.183***
	(0.040)	(0.001)	(0.001)	(0.001)	(0.004)	(0.010)
Title length	-0.053***	-0.002***	-0.000***	-0.000***	-0.004***	-0.012***
	(0.005)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Non-alphanumeric characters	4.431***	0.173***	0.141***	0.158***	0.409***	1.198***
	(0.267)	(0.008)	(0.005)	(0.005)	(0.028)	(0.070)
Number of authors	2.302***	0.127***	0.093***	0.106***	0.383***	0.909***
	(0.466)	(0.003)	(0.019)	(0.021)	(0.078)	(0.187)
Constant	-36.119***	-0.503***	-0.327***	-0.204***	1.485***	2.072***
	(1.436)	(0.096)	(0.057)	(0.069)	(0.206)	(0.491)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Journal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	232,391	232,391	232,391	232,391	232,391	232,391

Notes: The table reports coefficients and corresponding standard errors in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 8: Results of regression models based on papers from medicine and life sciences - excluding unclear cases

	OLS Total citations	NBR Total citations	OLS Log citations	OLS IHS citations	OLS 3 year citations	OLS 5 year citations
Tripartite phrase	32.904***	0.191***	0.189***	0.193***	3.481***	9.380***
	(5.266)	(0.021)	(0.014)	(0.014)	(0.875)	(1.984)
Article age	10.434***	0.093***	0.080***	0.082***	-0.573***	-0.960***
	(0.598)	(0.002)	(0.002)	(0.001)	(0.170)	(0.322)
Number of pages	2.950	0.043***	0.007*	0.007***	0.918	1.679
	(1.812)	(0.004)	(0.004)	(0.000)	(0.610)	(1.105)
Title length	-0.291***	-0.002***	-0.001***	-0.001***	-0.022**	-0.057***
	(0.054)	(0.000)	(0.000)	(0.000)	(0.010)	(0.020)
Non-alphanumeric characters	33.257***	0.211***	0.148***	0.151***	3.890***	9.206***
	(5.170)	(0.013)	(0.009)	(0.009)	(0.495)	(1.158)
Number of authors	1.111***	0.008***	0.002***	0.002***	0.266***	0.566***
	(0.268)	(0.001)	(0.000)	(0.000)	(0.061)	(0.129)
Constant	-87.238***	2.520***	2.597***	3.219***	14.380**	25.903**
	(20.739)	(0.094)	(0.080)	(0.072)	(6.609)	(12.187)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Journal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	90,928	90,928	90,928	90,928	90,928	90,928

Notes: The table reports coefficients and corresponding standard errors in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 9: Results of regression models based on papers from medicine and life sciences including article quality scores - excluding unclear cases

	OLS Total citations	NBR Total citations	OLS Log citations	OLS IHS citations	OLS 3 year citations	OLS 5 year citations
Tripartite phrase	32.776***	0.191***	0.188***	0.193***	3.463***	9.342***
	(5.214)	(0.020)	(0.014)	(0.014)	(0.868)	(1.969)
Article age	11.303***	0.094***	0.082***	0.084***	-0.450***	-0.703**
	(0.603)	(0.002)	(0.002)	(0.001)	(0.169)	(0.319)
Number of pages	2.887	0.041***	0.007*	0.007***	0.909	1.660
	(1.765)	(0.004)	(0.004)	(0.000)	(0.604)	(1.093)
Title length	-0.270***	-0.002***	-0.001***	-0.001***	-0.019*	-0.051***
	(0.054)	(0.000)	(0.000)	(0.000)	(0.010)	(0.020)
Non-alphanumeric characters	30.818***	0.204***	0.142***	0.145***	3.544***	8.486***
	(5.172)	(0.013)	(0.009)	(0.008)	(0.495)	(1.158)
Number of authors	1.091***	0.008***	0.002***	0.002***	0.263***	0.560***
	(0.268)	(0.001)	(0.000)	(0.000)	(0.061)	(0.128)
Article Quality	39.426***	0.099***	0.093***	0.094***	5.606***	11.639***
	(2.774)	(0.003)	(0.002)	(0.002)	(0.301)	(0.679)
Constant	-162.911***	2.356***	2.419***	3.039***	3.620	3.564
	(21.240)	(0.095)	(0.079)	(0.071)	(6.524)	(12.058)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Journal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	90,928	90,928	90,928	90,928	90,928	90,928

Notes: The table reports coefficients and corresponding standard errors in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$